STATISTICS 15

A mathematics teacher is a mid-wife to ideas.
The first rule of discovery is to have brains and goodluck.
The second rule of discovery is to sit tight and wait till you get a braight idea.

- George Polya

15.1 Introduction

We have studied in class IX about the classification of given data into ungrouped data as well as grouped data frequency distributions. We have also learnt how to represent the data pictorially in the form of various graphs such as bar graphs, histogram with equal and unequal lengths of class and frequency polygons. We have studied the measures of central tendency, namely mean, median and mode of ungrouped data and grouped data. We have studied the concept of cumulative frequency curves called ogives.

15.2 Mean of grouped data

We know that the mean (or average) of ungrouped data is the sum of all the observations divided by the total number of observations. Recall that if $x_1, x_2, x_3, \ldots, x_k$ are observations with frequencies $f_1, f_2, f_3, \ldots, f_k$ respectively, then the sum of values of all the observations is $f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_kx_k$ and the total number of observations is $n = f_1 + f_2 + \ldots + f_k$.

So, the mean of the data is given by,

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_kx_k}{f_1 + f_2 + \ldots + f_k}$$

We have also symbolised the sum by the greek letter Σ (capital sigma). So,

$$\bar{x} = \frac{\sum_{i=1}^{k} f_ix_i}{\sum_{i=1}^{k} f_i}$$

When there is no doubt then $\bar{x}$ can be written as $\bar{x} = \frac{\Sigma f_ix_i}{\Sigma f_i}$ where i varies from 1 to k.

Example 1 : The marks obtained by 100 students of two classes in mathematics paper consisting of 100 marks are as follows :

Marks obtained (x_i)	15	20	25	32	35	45	50	60	70	77	80
Number of students (f_i)	2	3	7	4	10	12	9	8	6	8	11

Marks obtained (x_i)	85	90	92	95	99
Number of students (f_i)	9	4	2	3	2

Find the mean of the marks obtained by the students.

Solution : To find the mean we need the product of x_i with corresponding f_i. For that we prepare the following Table 15.1

Table 15.1

Marks obtained (x_i)	Number of students (f_i)	f_ix_i
15	2	30
20	3	60
25	7	175
32	4	128
35	10	350
45	12	540
50	9	450
60	8	480
70	6	420
77	8	616
80	11	880
85	9	765
90	4	360
92	2	184
95	3	285
99	2	198
	$\Sigma f_i = 100$	$\Sigma f_ix_i = 5921$

Now, $\overline{x} = \dfrac{\sum f_ix_i}{\sum f_i}$

$= \dfrac{5921}{100}$

$= 59.21$

Therefore the mean of the marks obtained by the students is 59.21.

In practice the data are large. So, for a meaningful study we have to convert the data into grouped data.

Now, let us convert the above data into grouped data by forming class-intervals of width 15 (because generally we take 6 to 8 classes and range of our data is 90. So we take class-interval as 15). Remember that, while allocating frequencies to each class-interval, a student achieving value equal to upper class-limit would be considered to be in the next class. For example, 7 students who have obtained 25 marks would be considered in the class 25-40 and not in the class 10-25. The grouped frequency distribution table is as Table 15.2.

Table 15.2

Class-interval	10-25	25-40	40-55	55-70	70-85	85-100
Number of students	5	21	21	8	25	20

Now we will see how to calculate mean in continuous frequency distribution. Here we use the formula $\overline{x} = \dfrac{\sum f_ix_i}{\sum f_i}$ where we take x_i as the mid-point of a class (which would serve as the representative of the whole class) and f_i as frequency of that class. It is assumed that the frequency of each class-interval is centred around its mid-point (or class-mark).

$$\text{Class mid-point} = \frac{\text{Upper class limit of the class} + \text{Lower class limit of the class}}{2}$$

For the above table the mid-point of the class 10-25 is $\frac{10+25}{2} = 17.5$

Similarly, we can find the mid-point of each class as shown in Table 15.3.

Table 15.3

Class-interval	Number of students (f_i)	Mid-point (x_i)	$f_i x_i$
10-25	5	17.5	87.5
25-40	21	32.5	682.5
40-55	21	47.5	997.5
55-70	8	62.5	500.0
70-85	25	77.5	1937.5
85-100	20	92.5	1850.0
	Total $\Sigma f_i = 100$		**$\Sigma f_i x_i = 6055$**

So, the mean of given data is given by $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{6055}{100} = 60.55$

This method of finding the mean is called **Direct Method.**

We can see that Table 15.1 and 15.3 are the same data and applying the same formula for the calculation of the mean but the results obtained are different. Why is it so, and which one is more accurate ? The difference in two values occurs because of the mid-point assumption in Table 15.3. 59.21 is the exact mean, while 60.55 is an approximate mean. It is assumed that the observation of a class is centred around mid-value.

Every time the values x_i s and f_is are not small. So when the numerical values of x_i and f_i are large, finding the product of x_i and f_i becomes tedious and time consuming also. So for such situation, let us think of a method of reducing these calculations. For this we cannot do anything for f_is but we can reduce x_i s to a smaller number so the calculation becomes easy. How can we do this ? Let us understand the method.

The first step is to choose one of the x as **assumed mean** and denote it by A. We may take as A an x_i which is the middle of $x_1, x_2, \ldots, x_n$. Let $d_i = x_i - A$.

[**Note :** In fact A can be any convenient number. There is no change in the proof given below.]

$$\bar{d} = \frac{\Sigma f_i d_i}{\Sigma f_i}$$

$$= \frac{\Sigma f_i (x_i - A)}{\Sigma f_i}$$

$$= \frac{\Sigma f_i x_i}{\Sigma f_i} - A \cdot \frac{\Sigma f_i}{\Sigma f_i}$$

$$= \frac{\Sigma f_i x_i}{\Sigma f_i} - A$$

$$\bar{d} = \bar{x} - A$$

$$\therefore \quad \bar{x} = A + \bar{d}$$

$$\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

Find $f_i d_i$ and $\Sigma f_i d_i$ as shown in Table 15.4. Let A = 62.5.

Table 15.4

Class-interval	Number of students (f_i)	Mid-point (x_i)	$d_i = x_i - A$	$f_i d_i$
10-25	05	17.5	–45	–225
25-40	21	32.5	–30	–630
40-55	21	47.5	–15	–315
55-70	08	62.5 = A	0	0
70-85	25	77.5	15	375
85-100	20	92.5	30	600
	$n = 100$			$\Sigma f_i d_i = -195$

Now, $\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$

Now, substituting the values, we get

$$\bar{x} = 62.5 + \frac{(-195)}{100}$$

$$= 62.5 - 1.95$$

$$= 60.55$$

Therefore, the mean of the marks obtained by the students is 60.55

The method discussed above is called the method of **Assumed Mean.**

Activity 1 : From Table 15.3, taking the value of A as 17.5, 32.5 and so on and calculate the mean. The mean determined in each case is the same, i.e. 60.55.

So, we can say that the mean does not depend upon the value of A. Therefore we can take value of A as any non-zero number, not necessary that it is one of the value of x_i s.

See that in Table 15.4, the values in column 4 are multiple of 15 (i.e. class-interval), so if we divide the value of column 4 by 15, we get a smaller number to multiply with f_i.

So, let $u_i = \frac{x_i - A}{c}$, where A is the assumed mean and c is the class-size.

Suppose $\bar{u} = \frac{\Sigma f_i u_i}{\Sigma f_i}$

Now, let us find the relation between $\bar{u}$ and $\bar{x}$.

We have $u_i = \frac{x_i - A}{c}$

So, $\bar{u} = \dfrac{\sum f_i\left(\frac{x_i - A}{c}\right)}{\sum f_i} = \dfrac{1}{c}\left[\dfrac{\sum f_i x_i - A\sum f_i}{\sum f_i}\right]$

$= \dfrac{1}{c}\left[\dfrac{\sum f_i x_i}{\sum f_i} - A\right]$

$= \dfrac{1}{c}[\bar{x} - A]$

$\therefore \quad c\bar{u} = \bar{x} - A$

$\therefore \quad \bar{x} = A + c \cdot \bar{u}$

$= A + c \cdot \dfrac{\sum f_i u_i}{\sum f_i}$

Now let us calculate u_i as shown in Table 15.5. Here c is 15.

Table 15.5

Class-interval	f_i	x_i	$u_i = \frac{x_i - A}{c}$	$f_i u_i$
10-25	05	17.5	–3	–15
25-40	21	32.5	–2	–42
40-55	21	47.5	–1	–21
55-70	08	62.5	0	0
70-85	25	77.5	1	25
85-100	20	92.5	2	40
	$\Sigma f_i = 100$			$\Sigma f_i u_i = -13$

$\bar{x} = A + c \cdot \bar{u}$

$= A + c \cdot \dfrac{\sum f_i u_i}{\sum f_i}$

$= 62.5 + 15\left(\dfrac{-13}{100}\right)$

$= 62.5 + 15(-0.13)$

$= 62.5 - 1.95$

$= 60.55$

The method discussed above is called the method of **Step-Deviation.**

We note that :

- the step-deviation method will be convenient to apply if the class length is constant.
- the mean obtained by all the three methods is same.
- the assumed mean method and step-deviation method simplify the calculations involved in the direct method.

Example 2 : Find the mean of the data given below by all the three methods :

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	8	3	20	3	4	8

Solution : Let A = 35 and c = 10

Class	f_i	x_i	$d_i = x_i - A$	$u_i = \frac{x_i - A}{c}$	$f_i x_i$	$f_i d_i$	$f_i u_i$
0-10	4	5	–30	–3	20	–120	–12
10-20	8	15	–20	–2	120	–160	–16
20-30	3	25	–10	–1	75	–30	–3
30-40	20	35 = A	0	0	700	0	0
40-50	3	45	10	1	135	30	3
50-60	4	55	20	2	220	80	8
60-70	8	65	30	3	520	240	24
	$\Sigma f_i = 50$				$\Sigma f_i x_i$ = 1790	$\Sigma f_i d_i$ = 40	$\Sigma f_i u_i$ = 4

Using the direct method, $\overline{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1790}{50} = 35.8$

Using the assumed mean method, $\overline{x} = A + \frac{\sum f_i d_i}{\sum f_i}$

$= 35 + \frac{40}{50} = 35 + 0.8 = 35.8$

Using the step-deviation method, $\overline{x} = A + \left(\frac{\sum f_i u_i}{\sum f_i}\right) \times c$

$= 35 + \left(\frac{4}{50}\right) \times 10$

$= 35 + 0.8 = 35.8$

Therefore the mean of the data is 35.8

Example 3 : The mean of the following frequency distribution is 16, find the missing frequency :

Class	0-4	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36
Frequency	6	8	17	23	16	15	–	4	3

Solution : Let the missing frequency be x, take A = 26, c = 4

Class	Frequency	x_i	$u_i = \frac{x_i - A}{c}$	$f_i u_i$
0-4	6	2	–6	–36
4-8	8	6	–5	–40
8-12	17	10	–4	–68
12-16	23	14	–3	–69
16-20	16	18	–2	–32
20-24	15	22	–1	–15
24-28	x	26 = A	0	0
28-32	4	30	1	4
32-36	3	34	2	6
	$\Sigma f_i = 92 + x$			$\Sigma f_i u_i = -250$

We take A = 26, so that the product $f_i u_i$ is zero when $f_i = x$.

$$\bar{x} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times c$$

$$16 = 26 + \left(\frac{-250}{92 + x}\right) \times 4$$

$$-10 = \frac{-1000}{92 + x}$$

$\therefore$ $92 + x = 100$

$\therefore$ $x = 8$

$\therefore$ The missing frequency is 8.

Example 4 : The distribution below shows the number of wickets taken by a bowler in one-day cricket matches. Find the mean number of wickets.

Number of wickets	20-60	60-100	100-150	150-250	250-350	350-450
Number of bowlers	7	5	16	12	2	3

Solution : Here the class size varies and x_i's are large. So we take A = 200 and c = 10 and apply the step-deviation method.

Number of wickets	Number of bowlers (f_i)	x_i	$d_i = x_i - 200$	$u_i = \frac{x_i - A}{c}$	$f_i u_i$
20-60	7	40	−160	−16	−112
60-100	5	80	−120	−12	−60
100-150	16	125	−75	−7.5	−120
150-250	12	200=A	0	0	0
250-350	2	300	100	10	20
350-450	3	400	200	20	60
	$\Sigma f_i = 45$				$\Sigma f_i u_i = -212$

$$\therefore \quad \bar{x} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times c$$

$$= 200 + \left(\frac{-212}{45}\right) \times 10$$

$$= 200 - 47.11$$

$$= 152.89$$

$\therefore$ The mean wickets taken by the bowler is 152.89.

Example 5 : The mean of the following frequency distribution of 125 observations is 22.12. Find the missing frequencies.

Class	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
Frequency	3	8	12	–	35	21	–	6	2

Solution : Let the missing frequencies for the classes 15-19 and 30-34 be respectively f_1 and f_2. Let A = 17 and c = 5

Class	Frequency	x_i	$u_i = \frac{x_i - A}{c}$	$f_i u_i$
0-4	3	2	−3	−9
5-9	8	7	−2	−16
10-14	12	12	−1	−12
15-19	f_1	17	0	0
20-24	35	22	1	35
25-29	21	27	2	42
30-34	f_2	32	3	$3f_2$
35-39	6	37	4	24
40-44	2	42	5	10
	$\Sigma f_i = 87 + f_1 + f_2$			$\Sigma f_i u_i = 74 + 3f_2$

Here the total number of observations is 125 and

$$\Sigma f_i = 87 + f_1 + f_2$$

$\therefore$ $125 = 87 + f_1 + f_2$

$\therefore$ $f_1 + f_2 = 38$ **(i)**

Now, $\bar{x} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times c$

$$22.12 = 17 + \left(\frac{74 + 3f_2}{125}\right) \times 5$$

$\therefore$ $5.12 = \frac{74 + 3f_2}{25}$

$\therefore$ $5.12 \times 25 = 74 + 3f_2$

$\therefore$ $128 - 74 = 3f_2$

$\therefore$ $54 = 3f_2$

$\therefore$ $f_2 = 18$. Also $f_1 + f_2 = 38$. So, $f_1 = 20$

$\therefore$ The missing frequencies are 20 and 18.

EXERCISE 15.1

1. Find the mean of the following frequency distribution :

Class	0-50	50-100	100-150	150-200	200-250	250-300	300-350
Frequency	10	15	30	20	15	8	2

2. Find the mean wage of 200 workers of a factory where wages are classified as follows :

Class	100-150	150-200	200-250	250-300	300-350	350-400	400-450	450-500	500-550
Frequency	4	8	14	42	50	40	32	6	4

3. **Marks obtained by 140 students of class X out of 50 in mathematics are given in the following distribution. Find the mean by method of assumed mean method :**

Class	0-10	10-20	20-30	30-40	40-50
Frequency	20	24	40	36	20

4. **Find the mean of the following frequency distribution by step-deviation method :**

Class	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	5	10	20	9	6	2

5. **Find the mean for the following frequency distribution :**

Class	1-5	6-10	11-15	16-20	21-25	26-30	31-35
Frequency	18	32	30	40	25	15	40

6. **A survey conducted by a student of B.B.A. for daily income of 600 families is as follows, find the mean income of a family :**

Income	200-299	300-399	400-499	500-599	600-699	700-799	800-899
Number of families	3	61	118	139	126	151	2

7. **The number of shares held by a person of various companies are as follows. Find the mean :**

Number of shares	100-200	200-300	300-400	400-500	500-600	600-700
Number of companies	5	3	3	6	2	1

8. **The mean of the following frequency distribution of 100 observations is 148. Find the missing frequencies f_1 and f_2 :**

Class	0-49	50-99	100-149	150-199	200-249	250-299	300-349
Frequency	10	15	f_1	20	15	f_2	2

9. **The table below gives the percentage of girls in higher secondary science stream of rural areas of various states of India. Find the mean percentage of girls by step-deviation method :**

Percentage of girls	15-25	25-35	35-45	45-55	55-65	65-75	75-85
Number of states	6	10	5	6	4	2	2

10. **The following distribution shows the number of out door patients in 64 hospitals as follows. If the mean is 18, find the missing frequencies f_1 and f_2 :**

Number of patients	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of hospitals	7	6	f_1	13	f_2	5	4

*

15.3 Mode of Grouped Data

Let us recall that the observation which is repeated most often in an ungrouped data is called the mode of the data. In this chapter we shall discuss the way of obtaining the mode of grouped data, denoted by Z.

Let us recall how to find mode of ungrouped data through the following example.

Example 6 : The wickets taken by a bowler in 10 one-day matches are as follows :
4, 5, 6, 3, 4, 0, 3, 2, 3, 5. Find the mode of the data.

Solution : Here 3 is the number of wickets taken by a bowler in maximum number of matches. i.e. 3 times. So the mode of this data is 3.

In grouped frequency distribution, it is not possible to determine the mode of the data by looking at the frequencies. Here, we can only locate a class with the largest frequency, called the **modal class**. The mode is a value inside the modal class and it is given by the formula :

$$Z = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$$

where l = lower boundary point of the modal class

c = size of class interval (assuming all class sizes to be equal)

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class.

f_2 = frequency of the class succeeding the modal class.

Let us use this formula in the following examples.

Example 7 : A survey conducted on 20 hostel students for their reading hours per day resulted in the following frequency table :

Number of reading hours	1–3	3–5	5–7	7–9	9–11
Number of hostel students	7	2	8	2	1

Find the mode of this data.

Solution : Here the maximum class freqency is 8 and the class corresponding to this frequency is 5–7. So, the modal class is 5–7.

∴ The lower limit of the modal class 5–7 is $l = 5$.

Class size $c = 2$ and frequency of the modal class $f_1 = 8$. Frequency of the class preceding the modal class $= f_0 = 2$ and frequency of the class suceeding the modal class $= f_2 = 2$.

Now let us substitute these values in the formula :

$$Z = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$$

$$= 5 + \left(\frac{8 - 2}{2 \times 8 - 2 - 2}\right) \times 2$$

$$= 5 + \frac{6}{12} \times 2$$

$$= 5 + 1 = 6$$

So, the mode of above data is 6.

Example 8 : The mark distribution of 30 students at mathematics examination in a class is as below :

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Number of students	05	21	21	08	25	20

Find the mode of this data.

Solution : Here the maximum class frequency is 25 and the class corresponding to this frequency is 70-85. So, the modal class is 70-85.

The lower limit l of modal class 70-85 = 70 and class size $c = 15$

Frequency of the modal class $f_1 = 25$

Frequency of the class preceding the modal class $= f_0 = 08$,

Frequency of the class succeeding the modal class $= f_2 = 20$.

Now, let us substitute there values in the formula :

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$$

$$= 70 + \left(\frac{25 - 8}{2 \times 25 - 8 - 20}\right) \times 15$$

$$= 70 + \frac{17 \times 15}{22}$$

$$= 70 + 11.59 = 81.59$$

So, the mode of the above data is 81.59.

EXERCISE 15.2

1. Find the mode for the following frequency distribution :

Class	4-8	8-12	12-16	16-20	20-24	24-28
Frequency	9	6	12	7	15	1

2. The data obtained for 100 shops for their daily profit per shop are as follows :

Daily profit per shop (in ₹)	0-100	100-200	200-300	300-400	400-500	500-600
Number of shops	12	18	27	20	17	6

Find the modal profit per shop.

3. Daily wages of 90 employees of a factory are as follows :

Daily wages (in ₹)	150-250	250-350	350-450	450-550	550-650	650-750	750-850	850-950
Number of employees	4	6	8	12	33	17	8	2

Find the modal wage of an employee.

4. Find the mode for the following data : (4 and 5)

Class	0-7	7-14	14-21	21-28	28-35	35-42	42-49	49-56
Frequency	26	31	35	42	82	71	54	19

5.

Class	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180
Frequency	11	14	18	21	31	27	12	11	10

6. The following data gives the information of life of 200 electric bulbs (in hours) as follows :

Life in hours	0-20	20-40	40-60	60-80	80-100	100-120
Number of electric bulbs	26	31	35	42	82	71

Find the modal life of the electric bulbs.

*

15.4 Median of Grouped Data (M)

We have seen the definition of median for ungrouped data in standard IX as : "After arranging the observations in ascending or descending order, the number which is obtained in the middle is called the median." **Also, if the number of observations n is odd, then $\left(\frac{n+1}{2}\right)^{th}$ observation is the median and if the number of observations n is even, then median $M = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2}$.**

Suppose we have to find the median of the following data, which shows the marks of 50 students in mathematics out of 50 marks :

Marks obtained	18	22	30	35	39	42	45	47
Number of Students	4	5	8	8	16	4	2	3

Here $n = 50$ which is even. The median will be the average of $\frac{n}{2}^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observation i.e. 25th and 26th observations. To find this observation we proceed as follows :

Table 15.6

Marks obtained	Number of students
18	4
less than or equal to 22	4 + 5 = 9
less than or equal to 30	9 + 8 = 17
less than or equal to 35	17 + 8 = 25
less than or equal to 39	25 + 16 = 41
less than or equal to 42	41 + 4 = 45
less than or equal to 45	45 + 2 = 47
less than or equal to 47	47 + 3 = 50

We have formed a column which shows the number of students getting marks less than or equal to a particular number. It is known as cumulative frequency column.

Table 15.7

Marks obtained	Number of students (f)	Cumulative frequency (cf)
18	4	4
22	5	9
30	8	17
35	8	25
39	16	41
42	4	45
45	2	47
47	3	50

From the above table, we see that 25th observation is 35
26th observation is 39

$\therefore$ Median $= \frac{35+39}{2} = 37$

Note : The column 1 and column 3 of table 15.7 form cumulative frequency table. The median 37 shows the information that 50 % students obtained marks less than 37 and remaining 50 % students obtained marks more than 37.

Now let us see how to obtain the median of a grouped data from the following.

Consider a grouped frequency distribution of marks obtained, out of 100, by 55 students, in a certain examination, as follows :

Table 15.8

Marks	Number of students
0-10	2
10-20	3
20-30	3
30-40	4
40-50	3
50-60	4
60-70	7
70-80	11
80-90	8
90-100	10

We can see that 2 students obtained marks between 0 and 10, 3 students obtained marks between 10 and 20 and so on. So number of students who obtained marks less than 30 is $2 + 3 + 3 = 8$. Therefore the cumulative frequency of class 20-30 is 8. Similarly, we can obtaine the cumulative frequency of each class as shown in Table 15.9 as follow :

Table 15.9

Marks obtained	Number of students (cumulative frequency)
Less than 10	2
Less than 20	$2 + 3 = 5$
Less than 30	$5 + 3 = 8$
Less than 40	$8 + 4 = 12$
Less than 50	$12 + 3 = 15$
Less than 60	$15 + 4 = 19$
Less than 70	$19 + 7 = 26$
Less than 80	$26 + 11 = 37$
Less than 90	$37 + 8 = 45$
Less than 100	$45 + 10 = 55$

The distribution given in Table 15.9 is called **cumulative frequency distribution of less than type.** Here 10, 20, 30,..., 100 are the upper limits of the respective class intervals.

Similarly, we can make the table for the number of students with scores, more than or equal to 0, more than or equal to 10, more than or equal to 20 and so on. From Table 15.8, we can see that all

55 students have obtained marks more than or equal to 0. Since 2 students obtained marks in the interval 0-10, there are 55 − 2 = 53 students getting more than or equal to 10 marks and so on, as shown in Table 15.10.

Table 15.10

Marks obtained	Number of students (cumulative frequency)
More than or equal to 0	55
More than or equal to 10	55 − 2 = 53
More than or equal to 20	53 − 3 = 50
More than or equal to 30	50 − 3 = 47
More than or equal to 40	47 − 4 = 43
More than or equal to 50	43 − 3 = 40
More than or equal to 60	40 − 4 = 36
More than or equal to 70	36 − 7 = 29
More than or equal to 80	29 − 11 = 18
More than or equal to 90	18 − 8 = 10

The above table is called **cumulative frequency distribution of more than type.** Here 0, 10, 20,.., 90 are the lower limits of the respective class intervals.

Now, to find the median of this grouped data, we will make a table showing cumulative frequency with class interval and frequency, as shown in Table 15.11.

Table 15.11

Marks	Number of students (f)	Cumulative frequency (cf)
0-10	2	2
10-20	3	5
20-30	3	8
30-40	4	12
40-50	3	15
50-60	4	19
60-70	7	26
70-80	11	37
80-90	8	45
90-100	10	55

Here in a grouped data, we are not able to find the middle observation by looking at the cumulative frequencies as the middle observation will be some value in a class interval. So, it is necessary to find the value inside a class which divides the whole distribution into the halves. Which class is this ?

To find this class, we find the cumulative frequency n of all the classes and find $\frac{n}{2}$. Now we find the class whose cumulative frequency is greater than $\frac{n}{2}$ and nearest to $\frac{n}{2}$ is called the **median class**. In the distribution above, $n = 55$. So $\frac{n}{2} = 27.5$. Now, 70 − 80 is the class whose cumulative frequency is 37 which is just greater than 27.5. Therefore, 70-80 is the **median class.**

[**Note :** The cumulative frequency is just greater than $\frac{n}{2}$ means the smallest cumulative frequency which is cumulative frequency greater than $\frac{n}{2}$.]

After finding the median class, we use the formula given below for calculation of the median.

Median (M) = $l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times c$

where l = lower boundary point of the median class
n = total number of observations (sum of the frequencies)
cf = cumulative frequency of the class preceding the median class.
f = the frequency of the median class
c = class size (assuming class sizes to be equal)

Substituting the values $\frac{n}{2} = \frac{55}{2} = 27.5$, $l = 70$, $cf = 26$, $f = 11$, $c = 10$ in the above formula, we get

$$\text{Median (M)} = 70 + \left(\frac{27.5 - 26}{11}\right) \times 10$$

$$= 70 + \left(\frac{1.5 \times 10}{11}\right) = 71.36$$

So, the half of the students have obtained marks less than 71.36 and the other half have obtained marks more than 71.36.

Example 9 : A survey regarding the weights (in *kg*) of 45 students of class X of a school was conducted and the following data was obtained :

Weight (in *kg*)	Number of students
20-25	2
25-30	5
30-35	8
35-40	10
40-45	7
45-50	10
50-55	3

Find the median weight.

Solution : Here the number of observations is 45.

$\therefore \quad n = 45$. So, $\frac{n}{2} = 22.5$

Now, we will prepare the table containing the cumulative frequency as below :

Weight (in kg)	Number of students (f)	Cumulative frequency (cf)
20-25	2	2
25-30	5	7
30-35	8	15
35-40	10	25
40-45	7	32
45-50	10	42
50-55	3	45

$\frac{n}{2} = 22.5$. This observation lies in the class 35-40. So the median class is 35-40.

So, $l = 35$, $cf = 15$, $f = 10$, $c = 5$

Using the formula $M = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times c$

$$= 35 + \left(\frac{22.5 - 15}{10}\right) \times 5$$

$$= 35 + \left(\frac{7.5 \times 5}{10}\right) = 38.75$$

So, the median weight is 38.75 *kg*.

This means that the 50 % students have more weight than 38.75 *kg* and other 50 % students have weight less than 38.75 *kg*.

Example 10 : The median of the following frequency distribution is 38.2. Find the value of x and y, where sum of the frequencies is 165.

Class	5-14	14-23	23-32	32-41	41-50	50-59	59-68
Frequency	5	11	x	53	y	16	10

Solution :

Class	Frequency	Cumulative frequency
5-14	5	5
14-23	11	16
23-32	x	$16 + x$
32-41	53	$69 + x$
41-50	y	$69 + x + y$
50-59	16	$85 + x + y$
59-68	10	$95 + x + y$

Solution : It is given that $n = 165$. So, $95 + x + y = 165$, i.e. $x + y = 70$

Also, the median is 38.2 which lies in the class 32-41.

So, median class is 32-41.

$\frac{n}{2} = \frac{165}{2} = 82.5$, $l = 32$, $cf = 16 + x$, $f = 53$, $c = 9$

Using the formula $M = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times c$

$$\therefore \quad 38.2 = 32 + \left(\frac{82.5 - 16 - x}{53}\right) \times 9$$

$$6.2 = \frac{66.5 - x}{53} \times 9$$

$$\therefore \quad \frac{6.2 \times 53}{9} = 66.5 - x$$

$$\therefore \quad 36.5 = 66.5 - x$$

$$\therefore \quad x = 30$$

but $x + y = 70$. So, $y = 40$

∴ The value of x and y are respectively 30 and 40.

Note : There is a relation between the measures of central tendency :

Mode (Z) = 3 Median (M) − 2 Mean ($\bar{x}$)

EXERCISE 15.3

1. **Find the median for the following :**

Value of variable	12	13	14	15	16	17	18	19	20
Frequency	7	10	15	18	20	10	9	8	3

2. **Find the median for the following frequency distribution :**

Class	4-8	8-12	12-16	16-20	20-24	24-28
Frequency	9	16	12	7	15	1

3. **Find the median from following frequency distribution :**

Class	0-100	100-200	200-300	300-400	400-500	500-600
Frequency	64	62	84	72	66	52

4. **The following frequency distribution represents the deposits (in thousand rupees) and the number of depositors in a bank. Find the median of the data :**

Deposit (₹ in thousand)	0-10	10-20	20-30	30-40	40-50	50-60
Number of depositors	1071	1245	150	171	131	8

5. **The median of the following frequency distribution is 38. Find the value of a and b if the sum of frequences is 400 :**

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	42	38	a	54	b	36	32

6. **The median of 230 observations of the following frequency distribution is 46. Find a and b :**

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	12	30	a	65	b	25	18

7. The following table gives the frequency distribution of marks scored by 50 students of class X in mathematics examination of 80 marks. Find the median of the data :

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	2	5	8	16	9	5	3	2

*

15.5 Graphical Representation of Cumulative Frequency Distribution

We know that 'one picture is better than thousand words.' In class IX, we have represented the data through bar graphs, histogram, frequency polygons. Let us now represent a cumulative frequency distribution graphically.

For example, let us consider the cumulative frequency distribution given in table 15.9.

Remember that 10, 20, 30,..., 100 are the upper limits of the class intervals. To represent the data of table 15.9 graphically, we represent the upper limits of the class intervals on X-axis and their corresponding cumulative frequencies on Y-axis, choosing a convenient scale. The scale may not be same on both the axes. Now, let us plot the points corresponding to the ordered pairs given by (upper limit, corresponding cumulative frequency). i.e. (10, 2), (20, 5), (30, 8), (40, 12), (50, 15), (60, 19), (70, 26), (80, 37), (90, 45), (100, 55) on a graph paper. By joining them by a free hand smooth curve (See figure 15.1), we get a curve.

The curve we obtain is called a **cumulative frequency curve** or an **Ogive** (of the less than type). (See figure 15.1)

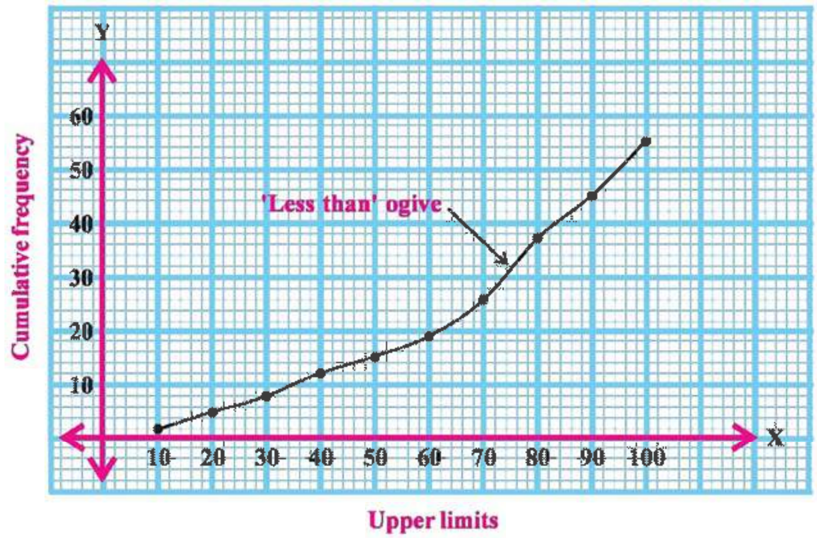

Figure 15.1

Now we draw the ogive (of more than type) of the cumulative frequency distribution in table 15.10.

Here, 0, 10, 20,...., 90 are the lower limits of the class intervals. To represent 'more than type' cumulative frequency curve, we plot the lower limits on X-axis and corresponding cumulative

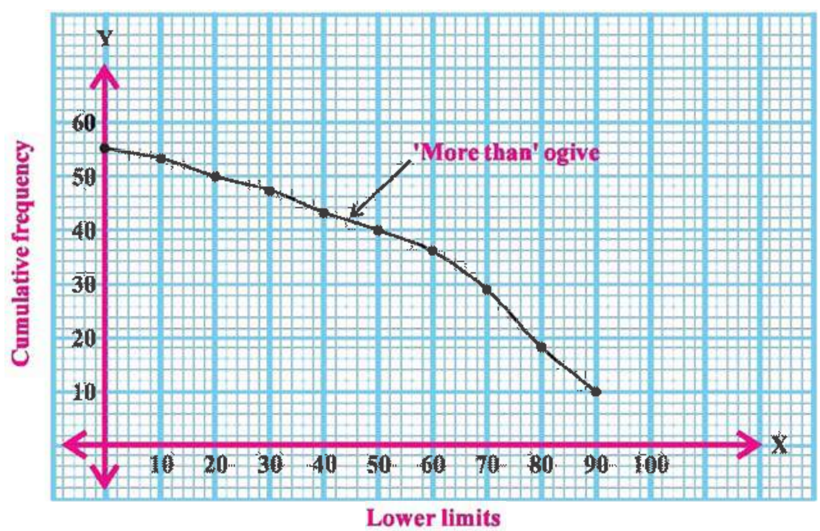

Figure 15.2

frequencies on Y-axis. Then we plot the points (lower limit, corresponding cumulative frequency), i.e. (0, 55), (10, 53), (20, 50), (30, 47), (40, 43), (50, 40), (60, 36), (70, 29), (80, 18), (90, 10) on a graph paper and join them by a free hand smooth curve. The curve we obtain is a **cumulative frequency curve** or **an ogive** (of more than type) (See figure 15.2)

In any way, are the ogives related to the median ?

One obvious way is to locate $\frac{n}{2} = \frac{55}{2} = 27.5$ on the Y-axis. From this point, draw a line parallel to X-axis intersecting the curve at a point. (See figure 15.3) From this point, draw a perpendicular to the X-axis. The point of intersection of this perpendicular with the X-axis determines the median of the data. (See figure 15.3)

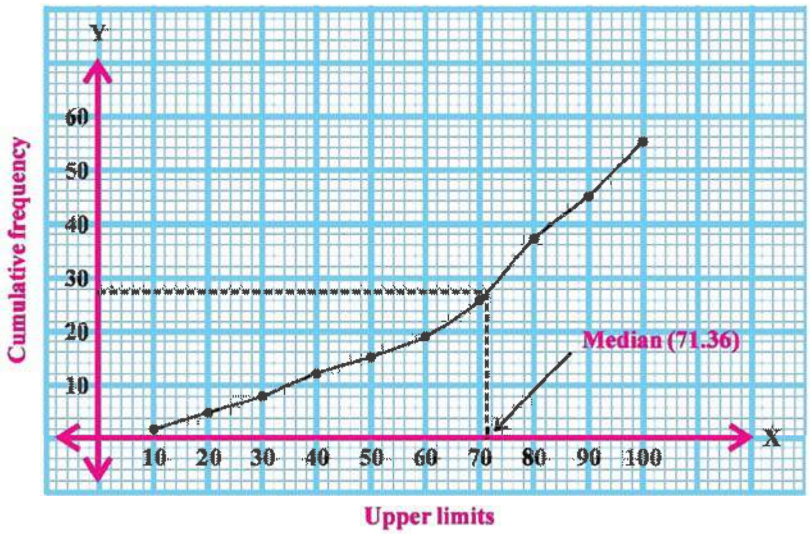

Figure 15.3

Another way of obtaining the median is as follows :

Draw both ogives (i.e. of less than type and of more than type) on the same graph-paper. The two ogives intersect each other at a point. From this point, if we draw perpendicular on the X-axis, the point at which it intersects the X-axis gives us the median (See figure 15.4)

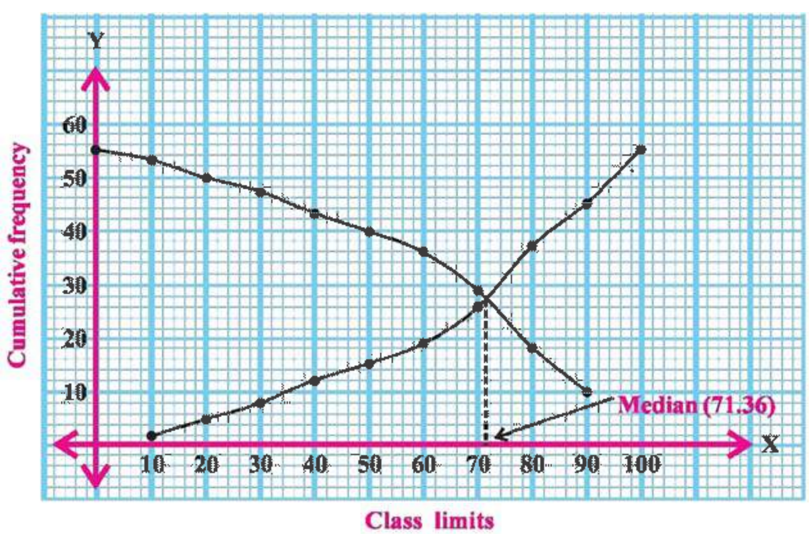

Figure 15.4

Example 11 : The annual income (in lakhs) of 30 officers in a factory gives rise to the following distribution :

Annual income (in lakh)	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Number of officers	2	9	3	6	4	4	2

Draw both ogives for the data above. Hence obtain the median annual income.

Solution :

Annual Income	Number of officers (f)	Cumulative frequency (cf)
5–10	2	2
10–15	9	11
15–20	3	14
20–25	6	20
25–30	4	24
30–35	4	28
35–40	2	30

First we draw the coordinate axes, with lower limits along the X-axis and cumulative frequency along the Y-axis. Then we plot the points (10, 2), (15, 11), (20, 14), (25, 20), (30, 24), (35, 28), (40, 30) for 'less than' ogive and (5, 30), (10, 28), (15, 19), (20, 16), (25, 10), (30, 6), (35, 2), for 'more than' ogive as shown in figure 15.5.

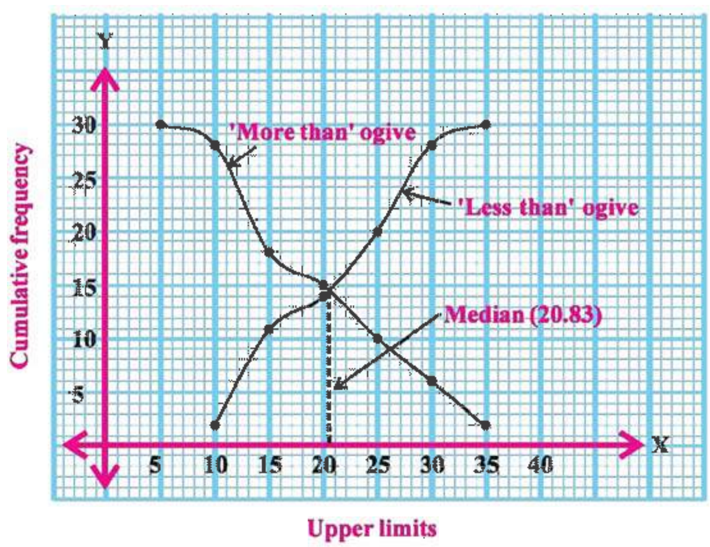

Figure 15.5

The x-coordinate of point of intersection is nearly 20.83, which is the median. It can also be verified by using the formula. Hence, the median annual income (in lakhs) is ₹ 20.83. (See figure 15.5)

EXERCISE 15

1. In a retail market, a fruit vendor was selling apples kept in packed boxes. These boxes contained varying number of apples. The following was the distribution of apples according to the number of boxes. Find the mean by the assumed mean number of apples kept in the box.

Nnmber of apples	50-53	53-56	56-59	59-62	62-65
Nnmber of boxes	20	150	115	95	20

2. **The daily expenditure of 50 hostel students are as follows :**

Daily expenditure (in ₹)	100-120	120-140	140-160	160-180	180-200
Number of students	12	14	8	6	10

Find the mean daily expenditure of the students of hostel using appropriate method.

3. **The mean of the following frequency distribution of 200 observations is 332. Find the value of x and y.**

Class	100-150	150-200	200-250	250-300	300-350	350-400	400-450	450-500	500-550
Frequency	4	8	x	42	50	y	32	6	4

4. **Find the mode of the following frequency distribution :**

Class	0-15	15-30	30-45	45-60	60-75	75-90	90-105
Frequency	8	16	23	57	33	23	13

5. **Find the mode of the following data :**

Class	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	12	17	28	23	7	8	5

6. **The mode of the following frequency distribution of 165 observations is 34.5. Find the value of a and b.**

Class	5-14	14-23	23-32	32-41	41-50	50-59	59-68
Freqnency	5	11	a	53	b	16	10

7. **Find the mode of the following frequency distribution :**

Class	1500-2000	2000-2500	2500-3000	3000-3500	3500-4000	4000-4500	4500-5000
Frequency	14	56	60	86	74	62	48

8. **Find the median of the following frequency distribution :**

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	9	11	15	24	19	9	8	5

9. **The median of the following data is 525. Find the value of x and y, if the sum of frequency is 100 :**

Class	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900	900-1000
Frequency	3	4	x	12	17	20	9	y	8	3

10. Select a proper option (a), (b), (c) or (d) from given options and write in the box given on the right so that the statement becomes correct :

(1) For some data, if Z = 25 and $\bar{x}$ = 25, then M = ...

(a) 25 (b) 75 (c) 50 (d) 0

(2) For some data Z − M = 2.5. If the mean of the data is 20, then Z = ...

(a) 21.25 (b) 22.75 (c) 23.75 (d) 22.25

(3) If $\bar{x}$ − Z = 3 and $\bar{x}$ + Z = 45, then M = ...

(a) 24 (b) 22 (c) 26 (d) 23

(4) If Z = 24, $\bar{x}$ = 18, then M = ...

(a) 10 (b) 20 (c) 30 (d) 40

(5) If M = 15, $\bar{x}$ = 10, then Z = ...

(a) 15 (b) 20 (c) 25 (d) 30

(6) If M = 22, Z = 16, then $\bar{x}$ = ...

(a) 22 (b) 25 (c) 32 (d) 66

(7) If $\bar{x}$ = 21.44 and Z = 19.13, then M = ...

(a) 21.10 (b) 19.67 (c) 20.10 (d) 20.67

(8) If M = 26, $\bar{x}$ = 36, then Z = ...

(a) 6 (b) 5 (c) 4 (d) 3

(9) The modal class of the frequency distribution given below is ......

Class	0-10	10-20	20-30	30-40	40-50
Frequency	7	15	13	17	10

(a) 10-20 (b) 20-30 (c) 30-40 (d) 40-50

(10) The cumulative frequency of class 20-30 of the frequency distribution given in (9) is ......

(a) 25 (b) 35 (c) 15 (d) 40

(11) The median class of the frequency distribution given in (9) is ......

(a) 40-50 (b) 30-40 (c) 20-30 (d) 10-20

*

Summary

In this chapter we have studied the following points :

1. The mean of the grouped data can be obtained by

(i) the direct method : $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$

(ii) the assumed mean method : $\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$

(iii) the step deviation method : $\bar{x} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times c$

assuming that the frequency of a class is centered at its mid-point.

2. The mode for the grouped data can be obtained by using the formula :

$$Z = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$$

where all symbols are in usual notations.

3. The cumulative frequency (cf) of a class is the frequency obtained by adding the frequencies of all the classes preceding the given class. The median for grouped data can be obtained by using the formula :

$$\text{Median (M)} = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times c$$

where all the symbols have their usual meaning.

Also $Z = 3M - 2\overline{x}$

4. Representing a cumulative frequency distribution graphically as a cumulative frequency curve or an ogive of 'less than type' and 'more than type' the median of the grouped data can be obtained graphically as the x-coordinate of the point of intersection of the two ogives.

Baudhayana, (fl. c. 800 BCE) was an Indian mathematician, who was most likely also a priest. He is noted as the author of the earliest Sulba Sutra—appendices to the Vedas giving rules for the construction of altars—called the Baudhayana Sulbasûtra, which contained several important mathematical results. He is older than the other famous mathematician Apastambha. He belongs to the Yajurveda school.

He is accredited with calculating the value of pi to some degree of precision, and with discovering what is now known as the Pythagorean theorem.

The sutras of Baudhayana :

The Shrautasutra

His shrauta sutras related to performing Vedic sacrifices has followers in some Smarta brahmanas (Iyers) and some Iyengars of Tamil Nadu, Kongu of Tamil nadu, Yajurvedis or Namboothiris of Kerala, Gurukkal brahmins, among others. The followers of this sutra follow a different method and do 24 Tila-tarpana, as Lord Krishna had done tarpana on the day before Amavasya; they call themselves Baudhayana Amavasya.